UNIVERSITY OF CALIFORNIA SAN DIEGO

Inducible Protein Depletion of SETD5, a Chromatin Regulator Implicated in Autism Spectrum Disorders

A Thesis submitted in partial satisfaction of the requirements

for the degree Master of Science

in

Biology

By

Mark Moyer

Committee in Charge:

Professor Alon Goren, Chair
Professor Heidi Cook-Andersen, Co-Chair
Professor Lorraine Pillus

2020

The Thesis of Mark Moyer is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ASD	Autism Spectrum Disorders
NDD	Neurodevelopmental Disorder
ID	Intellectual Disability
SFARI	Simon's Foundation Autism Research Initiative
CR	Chromatin Regulator
dTAG	Degradation Tag
SETD5	SET Domain Containing 5
FKBP12	FK506 Binding Protein 12
BSD	Blasticidin
PURO	Puromycin
CRIS-PITCh	Clustered Regularly Interspaced Short – Precise Integration into Target Chromosome
HA-tag	Human Influenza Hemagglutinin
WES	Whole Exome Sequencing
HDR	Homology Directed Repair
NGS	Next-Generation Sequencing

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Alon Goren, my Principal Investigator, Mentor, and Chair for all the support, encouragement, and wisdom over the years in guiding me to become a better scientist. Starting form the moment he accepted me into the lab, my experience has been fruitful and has inspired me to continue my work in to improving lives through a career in medicine.

Professor Heidi Cook-Andersen and Professor Lorraine Pillus, thank you for joining my committee even in these changing times with the pandemic at hand.

I would also like to acknowledge Drs. Laura Yelin-Bekerman and Sharona Shleizer-Burko for training me in my lab work, starting from when I had no experience.

To all the current members of the Goren Lab, I want to thank you for contributing to such an open and welcoming lab environment. Being part of a collective support system where we would all help, teach, and learn from each other was a memorable experience I hope to carry on in the future.

ABSTRACT OF THE THESIS

Inducible Protein Depletion of SETD5, a Chromatin Regulator Implicated in Autism Spectrum Disorders

By

Mark Moyer

Master of Science in Biology

University of California San Diego, 2020

Professor Alon Goren, Chair
Professor Heidi Cook-Andersen, Co-Chair

Autism Spectrum Disorder is one of the most genetically heterogenous neurodevelopmental disorders known making it difficult to characterize and understand its genetic basis. Efforts to screen and curate Autism genes have revealed an

overrepresentation of Chromatin Regulator genes, suggesting a significant role in the etiology of this condition. This paper seeks to establish a novel means of studying chromatin regulator genes involved in neurodevelopmental disorders by utilizing the dTAG inducible protein depletion system which would allow for specific, rapid, and reversible protein depletion. In this paper we construct the CRISPR plasmid vectors for the tagging of the Chromatin Regulator *SETD5* with the dTAG cassette for inducible protein depletion. One Cas9 vector targeting the *SETD5* Stop codon and two HDR vectors were produced, each with a different selection marker to allow for ease in producing homozygous clones in the future. Two confirmed heterozygous HEK293T cell lines were confirmed to contain the dTAG cassette through sequencing. Partial degradation of the SETD5 protein upon induction was demonstrated as expected with heterozygotes and blotting for the HA-tag further confirmed this result. Overall these plasmid vector constructs provide a means of *SETD5* tagging with an effective inducible protein system for further investigation into the effects of haploinsufficiency of *SETD5*. This framework is ideal for use in human embryonic stem cells which can recapitulate the neurodevelopmental stages during which haploinsufficiency has an effect. Therefore, it also lays a framework for the investigation of other genes implicated in neurodevelopment disorders.

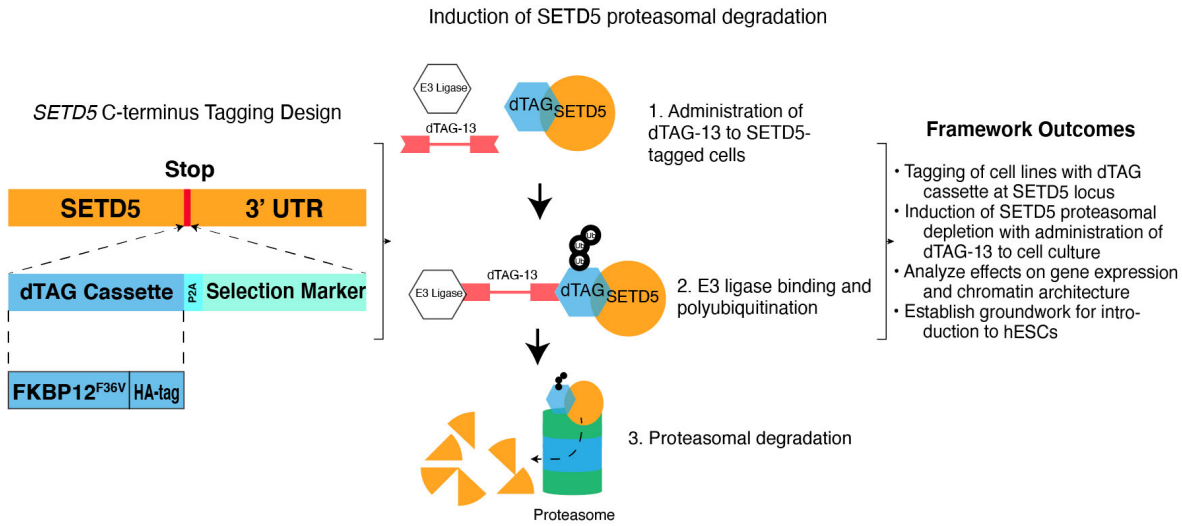


Figure 1: Graphical abstract of tagging of SETD5 and inducing depletion.

INTRODUCTION AND LITERATURE REVIEW

Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder (NDD) that affects how one understands, learns, and communicates leading to impeded social interactions [1]. This entails deficits in working memory, perception of emotions, and language development [2]. While it can be diagnosed at any age, symptoms typically begin to appear in the first two years of life [1]. An estimated 1 in 54 children in the U.S. are diagnosed with ASD [3]. Down syndrome, Fragile X syndrome, and Rett syndrome among other NDDs are also highly associated with ASD [1]. Individuals with ASD are observed to have a plethora of structural brain abnormalities observed through magnetic resonance imaging (MRI) [2]. Young children with Autism, 2-4 years in age, were found to have altered developmental trajectory with a higher total brain volume and increased cortical surface area [2]. At a cellular level, the prevailing hypothesis is that individuals with Autism display reduced GABA-ergic neuronal activity and thus have an excitation-inhibition imbalance [4]. In addition, many disfunctions in neurogenesis and migration, and the development of axons, dendrites, and synapses have been discovered [4].

The effort to understand the genetic basis of Autism has proven difficult as it is one of the most heterogeneous and genetically variable developmental disorders known. It was found that the 49% of Autism cases arise from inherited common variants of the reference genome whereas 3% are de novo copy number variants and 41% are unaccounted for [5]. Gene databases such as the Simons Foundation Autism Research Initiative (SFARI) database have curated Autism-related genes according to the strength of their association, yet how these hundreds of genes contribute to Autism is still

unknown [6]. Therefore, despite the advancements made with the analysis of whole genome studies, MRI scans, and cellular physiology, many of the genetic factors and how they contribute to Autism and other neurodevelopmental disorders are not well understood.

Throughout development, finely tuned processes dependent upon genetic and environmental factors contribute to the formation of the brain. Neuronal cell populations proliferate and differentiate, and then radial glia and neuronal cells migrate to their destinations [7]. Axons, dendritic spines, and synapses are developed to allow for electrochemical signaling, and these synapses are then pruned [7]. Oligodendrocytes myelinate axons and the spinal cord, and neural networks begin to form [7]. These complex processes are "epigenetically directed" meaning that they are dependent on the cells ability to change gene expression patterns according to a specific neurodevelopmental context [7]. These epigenetic changes in gene expression themselves are the restructuring of chromatin states within the cell to change which genes are closed off or open to transcription, which is determined by chromatin regulator proteins. It has been found that corticogenesis in humans, especially neuronal proliferation to synaptogensis, is prolonged and entails a vulnerability for genetic lesions [8]. Therefore, it makes sense that mutations Chromatin Regulators have been found to be one of the most common contributors to NDDs [8].

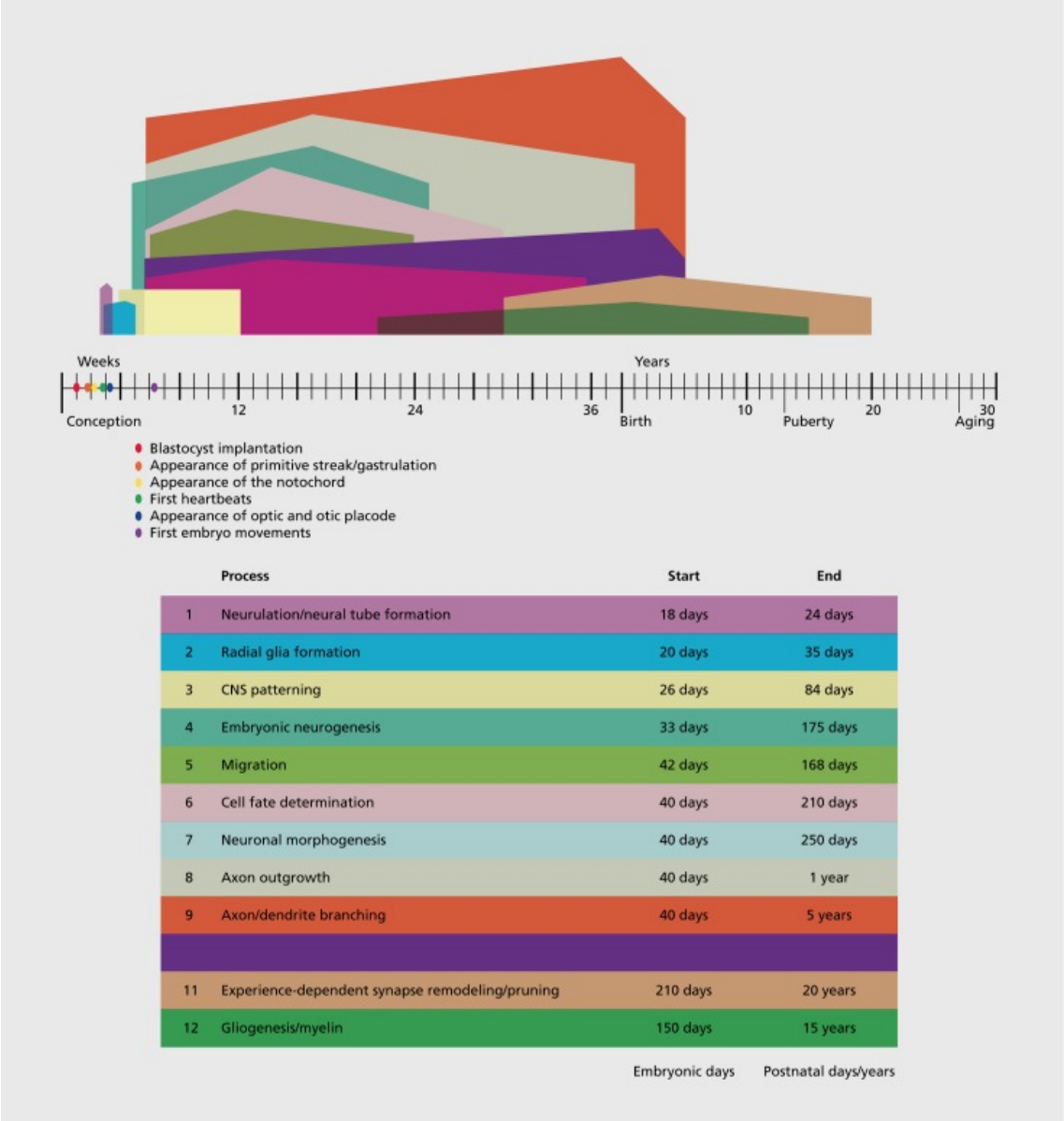


Figure 2: The human neurodevelopmental timeline. This figure conveys the stages of neurodevelopment with corresponding timelines relative to conception and birth. Conveyed in this figure is the prolonged nature of human corticogenesis. Reprinted from Pescosolido, 2012 [9].

Chromatin Regulators are highly critical in ASD etiology

Chromatin is a high-order structure composed of DNA, histones, and associated proteins that make up the chromosomes of eukaryotes [10]. Chromatin regulator proteins (CRs) include histone writers, erasers, readers and remodelers that can maintain or alter chromatin states to determine gene expression in the cell and are therefore of critical

importance in cell fate decisions [10]. For that reason, mutant CRs present in neurodevelopmental contexts could potentially contribute to significant downstream effects resulting in disorders such as ASD. This suggestion has been corroborated by several meta-analysis and review papers emphasizing the significance of Chromatin Regulators in ASD.

LaSalle et al. noted the recurrence and overrepresentation of chromatin remodelers, nucleosome packagers, and DNA methyltransferase in NDDs and ASD [11]. Specifically chromatin remodelers that demonstrate ATPase activity, involvement in demethylase activity, and methyltransferase activity are significantly overrepresented due to the fact that much of the early developmental genomic landscape is characterized by methylation of >80% of CpG islands and more importantly regulation of partially methylated domains (PMDs) which are highly enriched in neurodevelopmental genes [11]. Enrichment analysis of next-generation sequencing (NGS) results for several ASD cohorts of families also point towards chromatin regulators [12]. Additionally, many of Autism's associated diseases show Mendelian inheritance which contributes to the idea that rare de novo mutations in chromatin regulators lead to a loss of phenotypic robustness in neurodevelopment [12].

Our own analysis shows that chromatin regulators are 2.6 fold over-represented (P=1.62e-18) among suspected Autism genes curated by SFARI and highly ranked chromatin regulators have a 7.6 fold over representation (P=3.17e-15) (See Results: Figure 3 and 4). Overall, we can see that there is a significant involvement of chromatin regulators in the etiology of Autism and that this warrants a greater investigation into experimental frameworks that can be used to study it.

Means of studying genes involved in ASD

Given that in ASD many implicated genes see a disruption of function, the experimental framework for studying ASD should entail gene or protein perturbation to induce loss of function. The most common established methods of perturbation are CRISPR gene knockouts and RNA silencing. Since its inception, CRISPR gene knockout have been successful in inducing gene-specific heterozygous and homozygous knockout cell lines in a highly efficient manner [13]. On the other hand, RNA silencing is useful for rapid and reversible perturbation at the level of RNA without altering the genome, however it is known to have off target effects and can result in incomplete protein knockdown [14]. Therefore, while genetic knock-outs are precise, small molecular perturbations have the advantage of speed and reversibility. One additional method would be CRISPRi which utilizes an inactivated bacterial dCas9 and a customized guide RNA to bind the gene of interest, block the RNA polymerase, and suppress transcription [15]. The off-target effects of this system can be increased greatly depending on PAM sites available and the specificity of the 14 nucleotide long sgRNA-PAM sequence [15]. This short 14 nucleotide sequence could realistically be problematic for large genomes [15].

Inducible Protein Degradation systems offer an alternative to CRISPR knockouts and RNA silencing as it has the advantage of specificity for the gene/protein of interest while also inducible rapid degradation that can be reversed in recovery experiments. These inducible protein depletion systems can induce allele specific protein depletion within hours. Two well known protein degradation systems are the Auxin Inducible Degron (AID) system and the Degradation Tag (dTAG) system. The AID system works by tagging of the gene of interest with the AID tag in addition to knock-in of the plant-

OsTIR1 ubiquitin ligase that recognizes the AID and degrades the tagged protein product upon addition of auxin [16]. One caveat of this system is that the system may display chronic degradation even in the absence of auxin [16]. This leakiness has been corrected by the addition of an ARF cassette that also comes from the same plant cell machinery which prevents this chronic degradation; however that requires three knock-ins to implement this system [16].

The other inducible protein degradation system of interest in this project is the dTAG system. The dTAG system utilizes a mutant $FKBP12^{F36V}$ tag on a specific protein of interest which, when in the presence of the dTAG-13 chemical ligand, is recognized by endogenous ubiquitin proteasome cell machinery for immediate and allele-specific degradation [17, 18]. Specifically the $FKBP12^{F36V}$ tag facilitates binding to the Cereblon (CRBN) component of the E3 ubiquitin ligase complex which is endogenous to human cells [18]. This dTAG system displays no significant degradation of the tagged protein when the dTAG-13 ligand is not present indicating a more reliable system than the AID system [17]. In addition, one only needs to knock-in the tag because it utilizes endogenous cell machinery. The dTAG-13 ligand has demonstrated specificity only to the mutant $FKBP12^{F36V}$ and not to the endogenous wild-type FKBP12 [17]. One must also ensure that the tag does not alter the proteins functionality, which will ultimately be different from protein to protein [18]. This includes ensuring that the fusion chimera protein is not being degraded in the absence of dTAG-13 [18], and in our case comparing the expression pattern and chromatin binding patterns between tagged and wild type cells. The simplicity and efficiency of the dTAG system makes it the degradation system of choice for this project. With this system one can induce rapid and specific protein

depletion in a specific spatio-temporal context of brain development to better contrast differences between wild type and aberrational states. In addition, given its rapidity, it is unlikely in the case of heterozygote tagged cells, that the cell will have time to upregulate the wild-type allele [17].

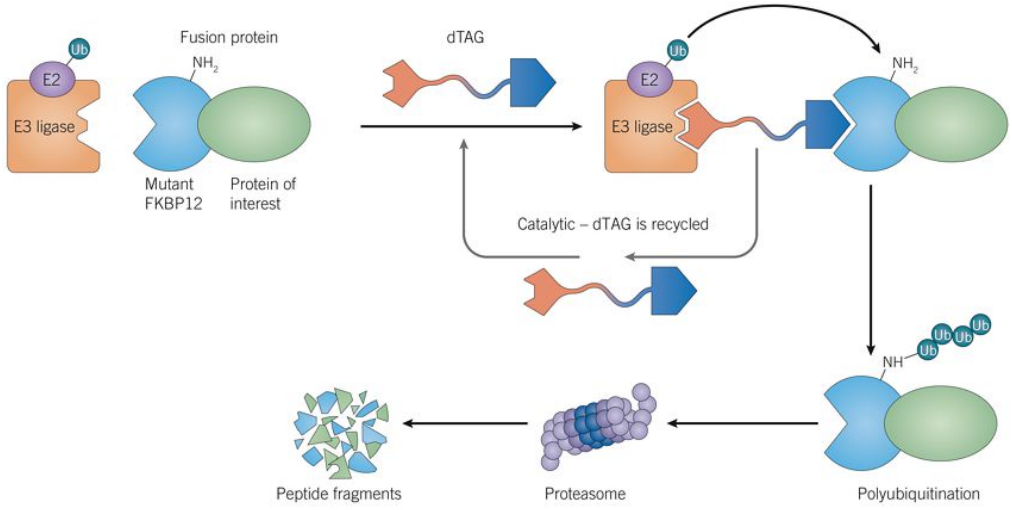


Figure 3: The dTAG molecular mechanism of action. A protein of interest contains the mutant FKBP12 from the dTAG cassette and upon administration of dTAG-13 to cell culture an endogenous E3 ubiquitin ligase binds the complex and ubiquitinates the mutant FKBP12, marking it for degradation by the proteasome. The dTAG-13 ligand is not degraded but recycled for continuous degradation. Reprinted from Nabet, 2018 [14].

Indications that specific neurodevelopmental stages, such as corticogensis, are convergence points for ASD mutations point towards the advantage of inducible degradation systems. Specifically, the advantage is having a pool of undifferentiated, gene-tagged cells that can be differentiated into many neurodevelopmental contexts to better recapitulate the appearance of de novo loss of function mutations. The fact that the tag can be used to deplete the protein both before inducing differentiation as well as after indicates that we can also decide whether to mimic pre-zygotic or post-zygotic mutations.

Candidate CRs highly implicated in ASD

The SFARI Gene Database ranked genes according to the strength of their association with ASD on a scale of 1 to 5, rank 1 being the strongest candidate genes [6]. The scoring of these autism genes is based entirely off of published, peer-reviewed scientific literature and is continuously updated to account for new data [6]. Central to their scoring methods is the use of de novo gene-disrupting mutations and genome wide significance as measures of increasing strength of association [6]. Genes can also be assigned the "Syndromic" label which means that they contain mutations associated with increased risk and are consistently linked to additional characteristics not required for an ASD diagnosis [6]. Our analysis of the 84 rank 1 and 2 genes, we observed that 24 are CRs, and 12 of those CRs are syndromic [6]. The CRs that fulfill both the criteria of being highly ranked and syndromic include: *ADNP*, *ARID1B*, *ASXL3*, *CHD2*, *CHD8*, *KMT2A*, *POGZ*, *SETD5*, *FOXP1*, *KMT2C*, *MECP2*, and *WAC* [6]. The first 8 are rank 1 genes, and the last 4 four rank 2 genes. All of these genes have been discovered to play a role in NDDs and intellectual disabilities. In addition, *KMT2A* and *KMT2C* also have a role in cancer cell proliferation meaning that our framework of studying CRs could potentially extend into cancer research.

ADNP (Activity-dependent neuroprotector homeobox): Mutations in *ADNP* have been shown to be associated with Helsmoortel-Van der Aa syndrome, Intellectual Disability (ID), altered facial characteristics, and congenital cardiac defects among others [19]. In a cohort of 78 individuals with disruptive *ADNP* mutations it was found that those with the p.Tyr719* mutation show the most severe phenotypes [19].

ARID1B (AT-rich interaction domain 1B): Mutations in *ARID1B* have been shown to be associated with Coffin-Siris Syndrome which causes developmental delays, altered facial characteristics, and hypotonia among other symptoms [20].

ASXL3 (ASXL transcriptional regulator 3): Mutations in *ASXL3* leads to Bainbridge-Ropers Syndrome which leads to ID, altered facial characteristics, and hypotonia among other symptoms [21]. In addition, autistic traits such as sleep disturbance, poor language development, and aggressive behavior are apparent with *ASXL3* mutations [21].

CHD2 (chromodomain helicase DNA binding protein 2): Mutations in *CHD2* have been how to cause onset of multiple types of seizures in the early stages of life, ID. ASD, and schizophrenia [22, 23].

CHD8 (chromodomain helicase DNA binding protein 8): Mutations in *CHD8* have been demonstrated to cause a particular subtype of ASD characterized by comorbidities such as macrocephaly and chronic GI issues among many other symptoms [24]. Mutations in its HELIC domain have been demonstrated to cause a more severe phenotype that entails delayed language development and overgrowth during puberty [25].

KMT2A (lysine methyltransferase 2A): Mutations in *KMT2A* have been demonstrated to cause Weidemann-Steiner Syndrome which is associated with developmental delays, ID, alter facial characteristics, and numerous other congenital malformations [26].

POGZ (pogo transposable element derived with ZNF domain): Mutations in *POGZ* have been demonstrated to cause ID, developmental delays, hypotonia, and altered facial characteristics [27].

SETD5 (SET domain containing 5): Mutations in *SETD5* have been demonstrated to cause ID, altered facial characteristics, and developmental delays [28, 29]. Whole exome sequencing of several large cohorts points towards de novo loss of function mutations in *SETD5* contributing to ASD [6]. To date there have been 25 reports of *SETD5* mutations in ID [6].

Table 1: Candidate Chromatin Regulators implicated in Autism. 12 CRs highly implicated in Autism as curated by the SFARI Gene Database [6]. Gene scores of 1 indicate a high confidence in it being an Autism gene and a score of 2 indicate a strong candidate Autsim gene. All 12 CRs are syndromic indicating a significant amount of "increased risk and consistently linked to additional characteristics" independent of an ASD diagnosis.

Gene Symbol	Gene Name	Chromosome	Support for Autism	Score	# Reports
ADNP	Activity-dependent neuroprotector homeobox	20q13.13	Rare Single Gene Mutation, Syndromic, Functional	1	34
ARID1B	AT-rich interaction domain 1B	6q25.3	Rare Single Gene Mutation, Syndromic	1	38
ASXL3	Additional sex combs like 3 (Drosophila)	18q12.1	Rare Single Gene Mutation, Syndromic	1	20
CHD2	Chromodomain helicase DNA binding protein 2	15q26.1	Rare Single Gene Mutation, Syndromic	1	31
CHD8	chromodomain helicase DNA binding protein 8	14q11.2	Rare Single Gene Mutation, Functional	1	42
FOXP1	forkhead box P1	3p13	Rare Single Gene Mutation, Syndromic, Genetic Association, Functional	2	44
KMT2A	Lysine (K)-specific methyltransferase 2A	11q23.3	Rare Single Gene Mutation, Syndromic	1	20
KMT2C	Lysine (K)-specific methyltransferase 2C	7q36.1	Rare Single Gene Mutation, Syndromic	2	15
MECP2	Methyl CpG binding protein 2	Xq28	Rare Single Gene Mutation, Syndromic, Functional	2	75
POGZ	Pogo transposable element with ZNF domain	1q21.3	Rare Single Gene Mutation, Syndromic	1	25
SETD5	SET domain containing 5	3p25.3	Rare Single Gene Mutation, Syndromic	1	23
WAC	WW domain containing adaptor with coiled-coil	10p12.1-p11.2	Rare Single Gene Mutation, Syndromic	2	17

Of these 12 candidate genes, we decided to use *SETD5*, "SET domain containing 5", to proceed forward with our study of Chromatin Regulators.

SETD5 is highly implicated in ASD chromatinopathies

SETD5 is a protein coding gene located on Chromosome 3 [6]. Other SET domain containing proteins are methyltransferases, and there is conflicting information concerning whether SETD5 is a methyltransferase. H3K9 histone methyltransferase activity was found in Pinheiro et al., yet in Deliu et al. it is claimed that SETD5 lacks methyltransferase activity and instead binds histone deacetylation HDAC3 complex [30, 31]. To date there are 25 reports on *SETD5* mutations, most of which seem to occur between positions 9,480,000-9,500,000 bp on chromosome 3[6, 28]. Like other chromatinopathies, mutations in SETD5 result in ASD, ID, altered facial characteristics, and developmental delays [28, 29].

From a cohort of 250 individuals with ID, 2 were found to have de novo mutations in SETD5 from Whole Exome Sequencing (WES) and an additional 4 were found to have de novo microdeletions through Chromasomal microarray [32] Recapitulation of the former two mutations in cells resulted in destruction of transcripts indicating that loss-of-function of SETD5 contributing towards the pathology [32]. In a cohort of 996 ID patients, 7 had loss of function mutations in the SETD5 gene [33]. These patients demonstrated phenotypes similar to those with 3p25 microdeletion syndrome which included altered cranial and facial characteristics, congenital heart defects and autism-like behaviors [33]. Whole exome sequencing of a 1.7 year old female with ID revealed a disrupting frameshift mutation in the SETD5 gene [34].

SETD5 haploinsufficient mice demonstrate hyperconnected cortical neurons with a reduction in excitatory neuronal synapses, along with a delay in network synchrony, and autism-like social behaviors [35]. *SETD5* homozygous knockout mice are lethal and

die between embryonic day 10.5 and 11.5 [36]. They displayed an inability to close the neural tube and had congenital heart defects [36]. At a molecular level, SETD5 was found to co-immunoprecipitate with histone deacetylase complexes and knockout of SETD5 led to increased acetylation of TSS sites in mESCs [36]. *SETD5* heterozygous knockout mice were demonstrated to have altered brain to body weight ratios and craniofacial abnormalities, differentially expressed genes tied to developmental defects, and behavioral and learning abnormalities with enhanced long-term potentiation which is tied to detriments in synaptic plasticity and ID [31]. *SETD5* heterozygous context fear condition mouse hippocampal tissues were found to have altered gene expression trajectories, especially in genes in PSD-encoding genes [31]. In addition, Deliu et al. found that SETD5 associates with Hdac3 and Paf1 complexes, occupies the TSS with Hdac3 and Pol II, and regulates Pol II occupancy of TSS in neurodevelopmental genes [31].

Project Outline

The goal of this project is to generate a means of studying *SETD5* function via targeted inducible degradation of the SETD5 protein. To do this we constructed and validated a plasmid vector for the tagging of the *SETD5* gene with the dTAG cassette. We first validate through sequence confirmation of the tag in plasmids and cell culture utilizing PCR and sanger sequencing. In addition, we demonstrate the inducible degradation of SETD5 works in cell cultures by performing dTAG experiments to induce depletion followed western blot to detect protein levels. This construct is a novel and particularly advantageous means of studying genes involved in neurodevelopmental disorders due to its specificity, rapidity, and reversibility. This project therefore provides

an alternative means of studying the effects of SETD5 haploinsufficiency, which would be continued by further experimentation. This system is particularly apt for use in human embryonic stem cells where introducing heterozygous or homozygous tagging of a SETD5 will allow study of haploinsufficiency and recovery in many different neurodevelopmental contexts.

METHODS

Plasmid Construction

In order to construct the Homology Directed Repair plasmids consisting of the dTAG cassette (linker- FKBP12^{F36V}-2xHA-P2A-BSD^{R}/Puro^{R}) flanked by 100+ bp homologous sequences, a Gibson assembly was performed. dTAG cassette sequences were amplified by PCR from the plasmids "pCRIS-PITChv2-dTAG-BSD" (Addgene #91795) and "pCRIS-PITChv2-dTAG-Puro" (Addgene #91796) with CloneAmp Hifi PCR Premix for high fidelity replication. The pUC19 plasmid backbone (Addgene #50005) was linearized with BamH1. Double strand G-block fragments were ordered from IDT containing the 100 base pair SETD5 homologies that will flank the dTAG cassette. Each G-block had 5' phosphorylation and contained 25 base pair homologies to the pUC19 site of insertion and the dTAG cassette. The InFusion protocol was used to perform blunt end DNA cloning and ligation with the linearized pUC19, dTAG cassette, and two G-blocks. This protocol was used to generate the two SETD5-dTAG HDR plasmids targeting the SETD5 C-terminus, one with the Blasticidin selection marker and one with the Puromycin selection marker.

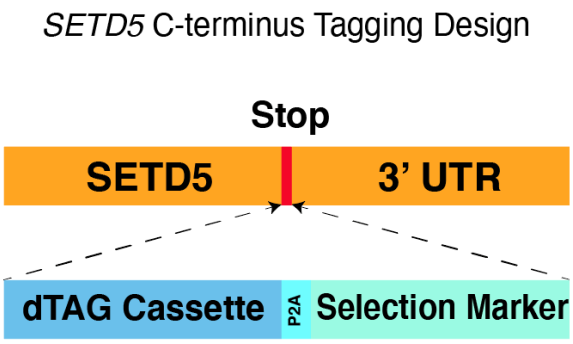


Figure 4: Knock-in of dTAG cassette into SETD5. The polycistronically encoded tag and selection marker portions of the dTAG cassette are flanked by homologies to the SETD5 protein encoding and 5' UTR regions. The endogenous stop codon is replaced by the stop codon encoded within the selection marker.

The Cas9-gRNA plasmid was formed from a pX330A-1x2 plasmid backbone (Addgene #58766). C-terminal guides were designed using the Benchling CRISPR Guide RNA design tool. Oligonucleotides from IDT for the gRNA sequence were ordered, phosphorylated with T4 PNK and ligation buffer, and annealed. The thermocycler program was 37 C for 30 minutes, 95 C for 5 minutes, then a ramp down to 25C at 5 C per minute. This product was diluted to 10nM. Next we performed a golden gate reaction with our px330A vector, diluted annealed oligos, NEB2 buffer, BbsI, ATP, BSA, water, and T7 DNA ligase. The thermocycler program was 37 C for 5 minutes and 20 C for 5 minutes for 20 cycles. This product was transformed into Stbl3 bacteria, colonies were picked, grown in LB, miniprepped and sequenced.

pUC19 Reconstruction

In order to reconstruct the pUC19 plasmids, which had a single base pair deletion in the SETD5 stop codon, we replicated the existing plasmids with the high fidelity polymerase found in the infusion kit.

Primers: BSD- Right Primer FOR: ATGAGCCCATAGCTGCTTC, Left Primer REV: TCCATCACTGTCCTTCACTATG; Puro- Right Primer FOR: ctaggccgagcatattgctgagg, Left Primer REV: TCCTTCGGGCACCTCGACGT

It was then ran on a gel, cut, and purified. We then ordered G-blocks for each of the pUC19 plasmids, one for puromycin resistance and one for blasticidin resistance.

BSD:5’-

GCTTCTCGATCTGCATCCTGGGATCAAAGCCATAGTGAAGGACAGTGATGGA

CAGCCGACGGCAGTTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTGGGA
GGGCTGAggcttctggatttgggcaaacagaactgaatgagcccatagctgcttccttccag-3’

Puromycin:5’-

GGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCAC
CTGGTGCATGACCCGCAAGCCCGGTGCCTGAggcttctggatttgggcaaacagaactgaatgag
cccatagctgcttccttccagctgcctctggaacctaggccgagcatattgctgaggaac-3’

These G-blocks contained the sequence upstream and downstream of the SETD5 stop codon and excluded the SETD5 stop codon. Therefore, we only required the stop codon of the antibiotic resistance genes and thus eliminated the frameshift deletion. We then performed an Infusion reaction, and 10 colonies were picked, grown in LB, and DNA was sent for sequencing. 4 Blasticidin colonies (bGG 178, 179, 180, 181) and 4 puromycin colonies (bGG 174, 175, 176, and 177) were found to have the correct sequences. Ultimately the chosen plasmids were bGG 176 (pUC19 SETD5 FKBP-V puro #4 reconstruction) and bGG 180 (pUC19 SETD5 FKBP-V BSD #4 reconstruction).

Cell Culture

HEK293T cells were grown in DMEM solution (2mM glutamine, 10%HI-FBS, 1% Anti-mycotic Anti-biotic) on standard tissue culture dishes or 6-well plates. Plates were stored in 37 ºC incubators with 5% CO_2.

Transfections

HEK293T cells were transfected with a Lipofectamine 3000 kit (Invitrogen #L3000015). A total amount of 3.4μg of plasmid was used for each transfection well, where the ratio of the Cas9-gRNA plasmid to the HDR plasmid was 1:1. Since the Cas9-gRNA plasmid is twice the length of the HDR plasmid, twice the weight of the HDR

plasmid was used. Therefore, 2.26µg of Cas9-gRNA plasmid and 1.13µg of HDR plasmid was used for each transfection well.

After 24 hours of incubation in transfection media, the media was switched back to DMEM. 72 hours after transfection, media containing antibiotics were used for 14 days. Cells transfected with the dTAG-Puromycin condition were subjected to 1 µg/ml puromycin and those with the dTAG-Blasticidin condition were subjected to 3 µg/ml blasticidin. As the wells cleared up by the 7-10 days of antibiotic treatment, colonies were picked and grown such that there would be cells used for DNA extraction and some frozen in Bambanker serum-free freezing media (Fisher Scientific #302-14681) at -80ºC for future use if needed.

Screening

DNA was extracted using our lysis buffer (10 mM Tris pH 7.5, 0.05% SDS, 25ug/ml proteinase K). The lysates were incubated at 37ºC for 1 hour and then heat inactivate at 80ºC for 30 min. PCR was performed using GoTaq Green MasterMix (Promega PRM7123), and 50 ng of DNA was used for template. In the case of the nested PCR reactions 1 µl of the previous reaction was used as template. The final primer concentration was 0.4 µM for each primer. Reactions were ran on a 1% agarose gel, and bands were extracted with Zymoclean Gel DNA Recovery kit (D4002).

Table 2: List of Primers. Primer sequences, with lab codes, and respective uses.

Primer Use	Code	Primer Sequence
Cassette Amplification/Confirm pUC19 sequences	GG1130	F: 5' – GGAGTGCAGGTGGAAACCAT – 3'
	GG1131	R: 5' – CCCGCAAGCCCGGTGCCTGA – 3'
	GG1132	R: 5' – GTTATGTGTGGGAGGGCTGA – 3'
Confirm pX330A sequences	GG1128	F: 5' – CACGGACTCAGACGGGACTTTCCTA – 3'
	GG1129	R: 5' – aaactaggaaagtcccgtctgagtc – 3'
pUC19 Reconstruction	GG1274	F: 5'-ATGAGCCCATAGCTGCTTC – 3'
	GG1275	R: 5' – TCCATCACTGTCCTTCACTATG – 3'
	GG1276	F: 5' – ctaggccgagcatattgctgagg – 3'
	GG1277	R: 5' – TCCTTCGGGCACCTCGACGT – 3'
Clonal Screen (1st half: Left homology arm - Selection marker)	GG1145	F: 5' – tgcccagtgctgggcagtcagc – 3'
	GG1148	R: 5' – TGAAGCCGAGCCGCTCGTAGA – 3'
	GG1147	R: 5' – AATTCACGAATCCCAACTGCC – 3'
Clonal Screen/Sequencing Primers Nested (1st half: Left homology arm - Selection marker)	GG1142	F: 5' – caggcctccagggtatctgcg – 3'
	GG1143	R: 5' - TTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGC – 3'
	GG1144	R: 5' – CTGTCCATCACTGTCCTTCAC – 3'
Clonal Screen/Sequencing Primers (2nd half: FKBP-V - Right homology arm)	GG1154	F: 5' – ATGCCACTCTCGTCTTCGAT – 3'
	GG1179	R: 5' – gactacgggctacaggcaag – 3'

Sequencing

Using the PCR primers, gel-extracted PCR fragments were sequenced with Sanger sequencing through Eton Biosciences. Primers sent in had a concentration of 10μM and PCR fragments were >10ng/μl.

Western Blot

Once incubations were completed we sought to preserve cells without lysing as that had proved to lead to degradation in protein amounts. We scraped cells with PBS, centrifuged the suspension at 320g for 5 min, and resuspended the cell pellet in 200μl PBS and froze at -80ºC. On the day of starting a western blot, pellets were thawed and lysed with a NP-40 Lysis Buffer (50mM Tris-HCl ph 8.5, 150mM NaCl, 1% NP-40,

cOmplete EDTA-free Protease Inhibitor 1X [Roche 11873580001], PhosSTOP 1X [Roche 4906845001], Benzonase 1:500). Afterwards the lysates were sonicated for ~20s on ice at 25 kHz, and centrifuged at 12,000 RPM at 4ºC for 20 minutes. Afterwards we ran a protein BCA assay (Fisher Scientific PI23227) to determine the protein concentrations. Protein amounts per lane were normalized to 15µg of protein. 6X Laemmli Buffer (375 mM Tris-HCl, 10% SDS, 7.5% glycerol, 0.03% bromophenol blue, 9% beta-mercaptoethanol) and remaining NP-40 buffer were added to samples to bring the total volume to 20 µl per lane.

Samples were boiled at 95ºC for 5 minutes. 7.5% precast polyacrylamide gel (BioRad #4568025) was loaded into a Bio-rad Mini Protean Tetra cell system and filled with running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS). 20 µl of sample was added into each lane, and Precision Plus Dual Color Standards (Bio-Rad 1610374) was used as a marker. Proteins were transferred on Nitrocellulose (0.45 µm, Invitrogen LC2001). The transfer system consisted of a 1X transfer buffer (25mM tris, 125 mM Glycine, 20% methanol pH 8.3) and was placed in a bucket filled with ice in the cold room. The transfer was 70 minutes at 120V. Afterwards the membranes were blocked for 1-1.5 hours at room temperatures in either TBS-T 5% milk or TBS-T 5% BSA.

Table 3: Antibodies and Western Blot conditions. Antibodies used for western blot with corresponding blocking conditions, primary solution conditions, and secondary solution conditions.

Antibody	Blocking	Dilution	Manufacturer	Catalog	Secondary Condition
BRD4	5% BSA in TBST	1:1000 5% milk in TBST	Bethyl	A700-004	1:3000 anti-rabbit secondary in TBST
Vinculin	5% milk in TBST	1:10,000 5% milk in TBST	Abcam	ab129002	1:5000 anti-rabbit secondary in TBST+5%milk
SETD5	5% milk in TBST	1:1000 5% milk in TBST	Invitrogen	PA5-57318	1:3000 anti-rabbit secondary in TBST
HA-tag	5% milk in TBST	1:1000 5% milk in TBST	Cell Signaling	#2367	1:2000 anti-mouse secondary in TBST

Antibodies used included: Cell Signaling HA-tag (6E2) Mouse mAb #2367, Bethyl BRD4 Recombinant Monoclonal Antibody [BL-149-2H5], A700-004, Abcam Recombinant Anti-Vinculin antibody [EPR8185] (ab129002), Invitrogen Life Technologies SETD5 Antibody (PA5-57318). Membranes were incubated in primary antibody solutions overnight in the cold room at 4ºC. Membranes were washed three times with TBST, incubated with secondary antibody solutions for 1 hour, and washed for another three times with TBST. Secondary antibodies included: Cell Signaling anti-mouse antibody (#7076), and Cell Signaling anti-rabbit antibody (#7074). Secondary antibody solutions were also composed of Precision Protein StrepTactin-HRP Conjugate (Bio-rad #1610381) for marker visualization at a 1:10,000 dilution. Proteins were visualized with the Thermo Scientific Chemiluminescent Nucleic Acid Detection Module Kit (ECL) (#89880) in a Bio-Rad Chemi-Doc XRS+ system.

RESULTS

A Statistical Analysis of CRs Illustrates Significant Role in Etiology of ASD

At the outset of this project, we sought to understand if there was a correlation between Chromatin Regulators and ASD-linked genes. Therefore, we performed the following statistical analysis that utilizes the database information that was provided at the time in 2019. The Simon's Foundation Autism Research Initiative (SFARI) curates genes from ASD patients and ranks them according to the strength of this association. The SFARI Gene Database listed 1019 genes suspected to be involved in ASD [6]. Rank 1 and 2 genes have the strongest association and rank 5 genes have the weakest association with ASD [6]. The Human Genome Organization (HUGO) listed 19194 genes in the human genome, and Epifactors listed 720 CR's in humans (~3.8% of all genes) [37, 38]. A statistical analysis via hypergeometric tests using these respective databases revealed an over-enrichment of CR's with respect to ASD. The hypergeometric test revealed that CR's make up 99 genes (~10%) of all 1019 Autism-linked genes which is a 2.6 fold over-enrichment (P=1.62e-18) with respect to what proportion of the genome they make up. They also make up 24 genes (~29%) of all 84 highly ranked (Rank 1 and 2) Autism-linked genes which is a 7.6 fold over-enrichment (P=3.17e-15). Overall, these results suggest CRs play a significant role in the etiology of ASD.

Hypergeometric Test:

		CR +	CR -
SFARI	+	Both CR and SFARI genes: 99 (a)	SFARI gene but not CR: 1019-99=920 (b)
	-	CR but not SFARI: 720-99=621 (c)	Neither SFARI nor CR: 19194 – (720-99+1019)=17554 (d)

Check: 99+920+621+17554=19194

k=99; s=1019; M=720; N=19194
N=population=a+b+c+d
M=successes in the population=a+c
s=sample=a+b
k=successes in the sample=a

Figure 5: Hypergeometric test of chromatin regulator overrepresentation in autism genes. Here sample is defined as all Autism genes as curated by SFARI Gene Database, whereas successes in the sample are the chromatin regulators, according to the Epifactors database, that are also Autism genes. The results show that CRs are **over enriched 2.59 fold** compared to expectations (p-value = 1.623036210970097e-18).

Hypergeometric Test:

		CR +	CR -
SFARI	+	Both CR and SFARI genes (Rank 1&2): 24 (a)	SFARI gene (Rank 1&2) but not CR: 84-24=60 (b)
	-	CR but not SFARI (rank 1&2): 720-24=696 (c)	Neither SFARI (Rank 1&2) nor CR: 19194 – (720-24+84)=18414 (d)

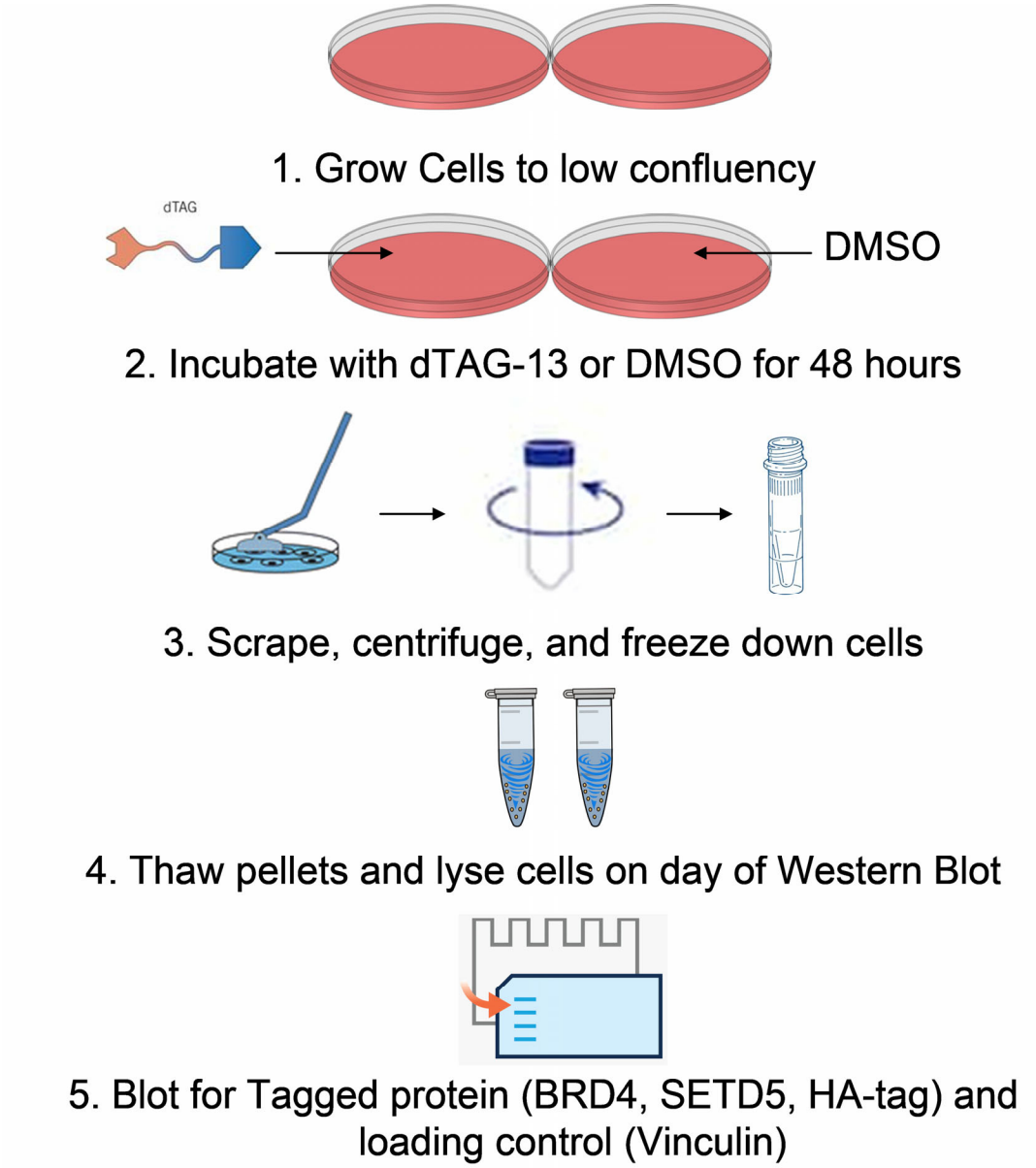
Check: 24+60+696+18414=19194

k=24; s=84; M=720; N=19194

Figure 6: Hypergeometric test of chromatin regulator overrepresentation in highly ranked autism genes. Here sample is defined as all highly ranked Autism genes as curated by SFARI Gene Database, whereas successes in the sample are the chromatin regulators, according to the Epifactors database, that are also highly ranked Autism genes. The results show that CRs are **over enriched 7.62 fold** compared to expectations (p-value = **3.171858239620918e-15**).

dTAG System Validation Demonstrates Complete Protein Depletion

We next sought to validate the efficacy and efficiency of the dTAG system in a preliminary cell line of homozygous BRD4-tagged HEK293T cells provided by the authors of the original dTAG study [17]. We ran dTAG experiments with two different concentrations of dTAG-13 in plates of *BRD4*-tagged and HEK293T cells with incubation times of 4 hours. Cell lysates were then used for western blot.



Scheme 1: dTAG Experiment Schematic.

We found that at both 250 nM and 500 nM dTAG -13, there was complete or near complete degradation of the BRD4 protein. Changing the loading amount made little difference as there were similar levels of depletion between 15ug protein and 30ug protein loading conditions. There was also no significant degradation of the BRD4 protein in the untreated and DMSO vehicle treated *BRD4*-tagged cells on par with the HEK293T wild type cells indicating that the tagged protein was stable and not being constitutively degraded by the endogenous cell proteasome machinery in the absence of the dTAG-13 ligand. These results suggest that the dTAG system is effective in inducing protein depletion and this process is specific to addition of the dTAG-13 ligand.

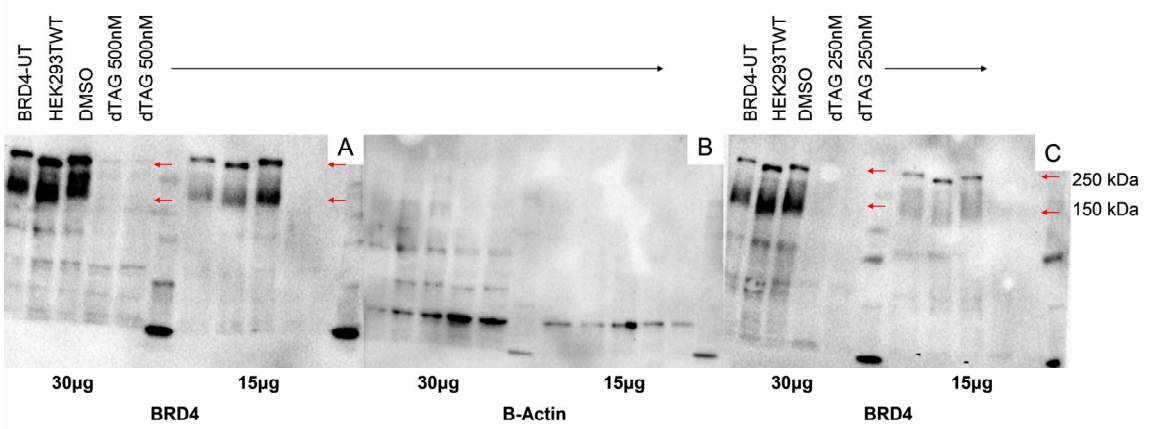


Figure 7: dTAG validation western blot. Blot for BRD4 on untreated BRD4 cells ("BRD4-UT"), wild type HEK293T cells, and BRD4 tagged cells either treated with DMSO or dTAG-13 at 250nM or 500nM. Loading amounts were either 15ug or 30ug. (A) dTAG-13 500nM condition demonstrates full degradation of short and long BRD4 isoforms. Red arrows indicate areas of degradation of short and long isoforms. (B) B-Actin re-blot of the membrane in (A) demonstrates roughly equal loading amount across the same loading condition. (C) dTAG-13 250nM condition also demonstrates complete degradation in both loading conditions.

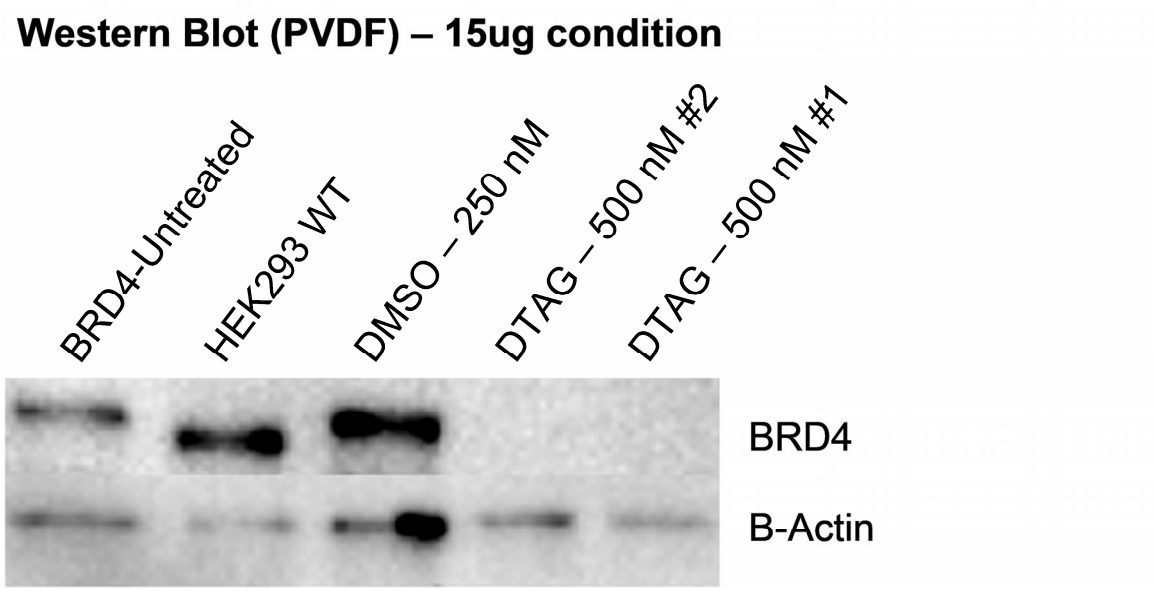


Figure 8: dTAG validation western blot summary.

Protein Depletion System Displays Reversibility

We have observed significant recovery of the BRD4 depletion 24-48 hours after passaging of dTAG-13 treated cells, however this recovery time could be earlier depending on how much recovery is desired and what gene is tagged. This reversibility was particularly enhanced by the transferal of the cells to a new plate as simple washouts were ineffective. Therefore, the dTAG-13 in its DMSO vehicle seems to stick to the cell culture plates quite well. The original authors of the dTAG paper also relayed this information to us and observed reversibility as well [17]. Overall, we can say that the dTAG system displays reversibility of protein depletion but that it is dependent upon transferal of cells to a new plate.

Initial SETD5-Tagging Design with PITCh-CRIS is Ineffective and Inefficient

For our initial design for tagging of the *SETD5* gene we used the method outlined in the dTAG paper which was the PITCh-CRIS system [17]. We decided to tag the C-terminus of the SETD5 gene right before the STOP codon as the available gRNA sites

were more favorable at that position. In the PITCh-CRIS system the homology regions flanking the dTAG cassette were around 25 base pairs, shorter than normal CRISPR HDR standards [39]. In PITCh-CRIS system, the cell purportedly uses alternative cell machinery from the well known CRISPR-Cas9 homology directed repair and microhomology regions flanking the insertion cassette would be utilized for microhomology-mediated end joining [17, 39].

After replicating our dTAG cassettes from the Addgene plasmids (Blasticidin #91795, Puromycin #91796) and ordering G-blocks with the microhomologies to the *SETD5* region, we used the InFusion kit to integrate them into our linearized pUC19 backbone with our microhomology regions flanking the cassette. After transfection of HEK293T cells with Lipofectamine 3000 and antibiotic treatments, and colonies were picked and grown. These colonies were screened for containing the dTAG cassette with PCR and sequencing. None were found to be positive. Upon contacting the original authors of the dTAG paper it was found that the PITCh-CRIS system is highly inefficient compared to CRISPR-Cas9 HDR as hundreds of colonies would need to be picked to find a positive clone. From this we learned that it was much more favorable to implement longer homology arms and utilize traditional CRISPR homology directed repair (HDR) for the purposes of this project. Thus, the PITCh-CRIS system is inefficient and not well suited to the needs of this project where one aspect of our goal is to provide the vectors and the framework for others to efficiently induce our knock-in cassette.

SETD5-Tagging Redesign with Traditional HDR Parameters is Successful

We then sought to redesign our tagging system to utilize traditional CRISPR-Cas9 HDR to tag the SETD5 gene. We redesigned the pUC19 HDR plasmids such that they

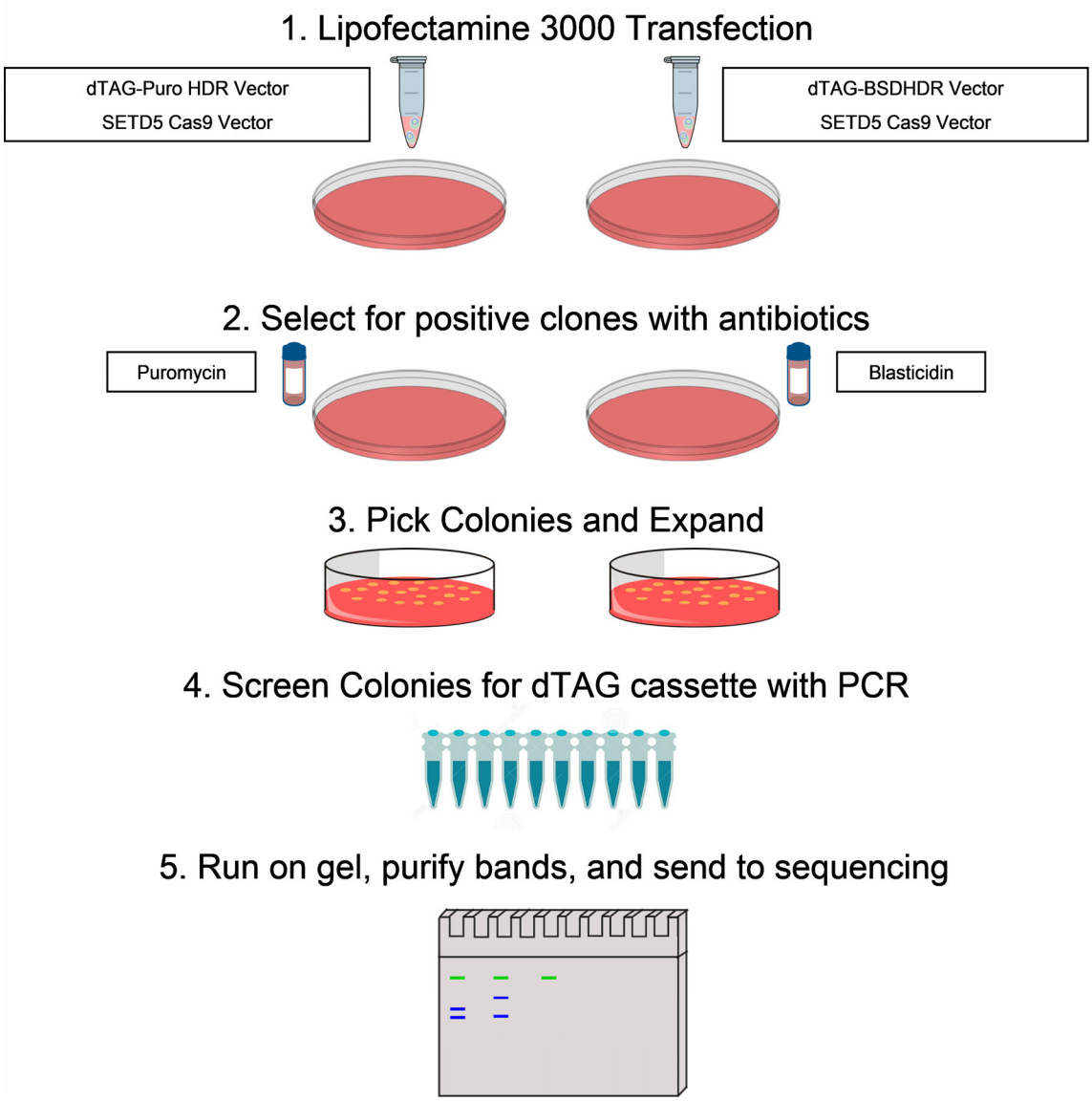
contained 100+ base pair homologies. This required complete reconstruction of the pUC19 plasmids from scratch from amplifying the dTAG cassettes from the original Addgene plasmids to performing an InFusion assembly with our dTAG cassette, G-blocks, and linearized pUC19 plasmid. The Cas9 plasmid was also made from scratch with an entirely new gRNA. The gRNA was integrated into a pX330A plasmid digested with BbsI by golden gate reaction. Once plasmid sequences were generated, transfections with the new plasmids were performed, and clones were picked. PCR screenings for colonies produced 3 puromycin clones and 3 blasticidin clones that appeared to have the cassette. These PCR sequences were cleaned from a gel, and sent to Eton Biosciences for sequencing. It was found that all the clones had the same 1 base pair deletion at the first base pair of the endogenous STOP codon. PCR and sequencing of the pUC19 plasmid sequences also found this same deletion which would result in a frameshift. The SETD5 STOP codon was also redundant as the antibiotic resistance genes contained STOP codons immediately prior. The pX330A plasmid targeting the SETD5 STOP codon (bGG 151) had the correct sequence however.

Therefore, the pUC19 plasmids were redesigned so that the stop codon of the gene was excluded. We linearized the two pUC19 plasmids by PCR with a high fidelity polymerase. We then ordered 170 base pair G-blocks that excluded the second stop codon, which contained the one base pair deletion, and used the InFusion reaction to produce plasmid colonies. Plasmids were then properly screened and sequenced and confirmed to have the correct sequence through Sanger sequencing from Eton Biosciences. Four SETD5-Puro colonies were confirmed to have the correct sequence (bGG 174, 175, 176, and 177) and four SETD5-BSD colonies were confirmed to have the

correct sequence (bGG 178, 179, 180, 181). The two plasmids of our choice were SETD5-Puro vector bGG 176 and SETD5-BSD vector bGG 180, which we used henceforth for our transfections.

Transfection, Screening, and Sequencing

Transfections on HEK293T were performed with Lipofectamine 3000, and our newly confirmed pUC19 HDR plasmids with our Cas9-gRNA plasmid. Initially colonies were picked and some 24 were successfully expanded. These colonies were screened for containing the dTAG cassette. 3 BSD colonies and 1 puro colony were confirmed to have the correct sequence from the end of the FKBP-V to the genome. Of those, 1 Blasticidin colony (D3) and 1 puromycin colony (A8) had the fully correct sequence from left homology region to cassette to right homology region. An additional 12 colonies were picked, expanded, and frozen for any future uses.



Scheme 2: Transfection and Clonal Screening Schematic.

Figure 9: dTAG cassette sequence of SETD5-Puro clone A8.

Figure 10: dTAG cassette sequence of SETD5-BSD clone D3.

SETD5 Clones Demonstrate Partial Degradation of SETD5

Next, we sought to validate whether degradation works with the SETD5 clones through a dTAG experiment. We incubated the four aforementioned *SETD5* heterozygote colonies and BRD4 tagged cells in dTAG-13 for 48 hours. Protein amounts were

visualized with Western Blot. We found that the dTAG-13 treated BRD4 tagged cells demonstrated complete depletion in comparison to the vehicle control as expected since they are homozygous for the tag. Next, we found that there was noticeable degradation of the SETD5 protein in 3 of the 4 clones (A8, D3, and D9), and little degradation of SETD5 in the fourth clone (E4). Among those 3 successful clones were the two clones that had their full sequence confirmed. The amount of degradation was not 100%, which was expected as they are heterozygous for the tag. Therefore, at most there should be ~50% depletion of the SETD5 protein. In addition, the SETD5 pattern demonstrated a double band pattern with the lower band showing more degradation than the higher band consistently. The DMSO controls showed no significant difference in SETD5 protein amount from the HEK293T control.

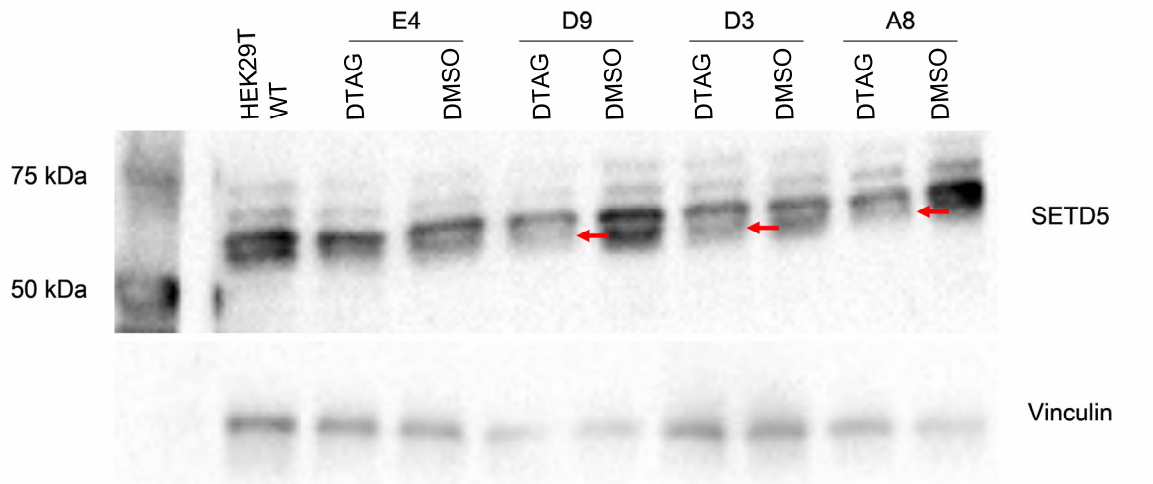


Figure 11: Degradation of SETD5 clones. SETD5 protein in 4 heterozygous SETD5-tagged cell lines, illustrating significant degradation of the SETD5 protein in clones A8, D3, and D9. Conditions incubated for 48 hours.

We also found that there was a significant difference in the confluency of the *BRD4*-tagged cells treated with dTAG-13 and the vehicle (DMSO). While both conditions were seeded at the same density from the same parental plate, by the end of

the incubation the DMSO-treated plates displayed a confluency of 80%, comparable to untreated cells. On the other hand, the dTAG-13 treated tagged cells displayed a significantly lesser confluency, not expanding past ~25% confluency, indicating that the protein depletion was rapid enough to induce effects on cell proliferation. These effects on confluency and morphology were not observed in dTAG-13 administration on HEK293T wild type cells as expected.

On the other hand, only some SETD5-tagged cell lines displayed differences in proliferation. Only in clones A8 and E4 did we see dTAG-13 decreasing proliferation in comparison to DMSO treatments. However, in all four clonal cell lines, administration of dTAG-13 altered the morphology of the HEK293T cells whereby cells appeared far more self-contained with the edges of the cultures having a smoother appearance. This could possibly mean that the fibroblast composition of the HEK293T cells was altered by the SETD5 depletion.

Overall, we found that our heterozygous SETD5-tagged cell lines displayed partial degradation of the SETD5 protein confirming that our knock-in effectively displays inducible protein depletion. In addition, depletion of either BRD4 or SETD5 results in less cell proliferation, and altered cell morphologies.

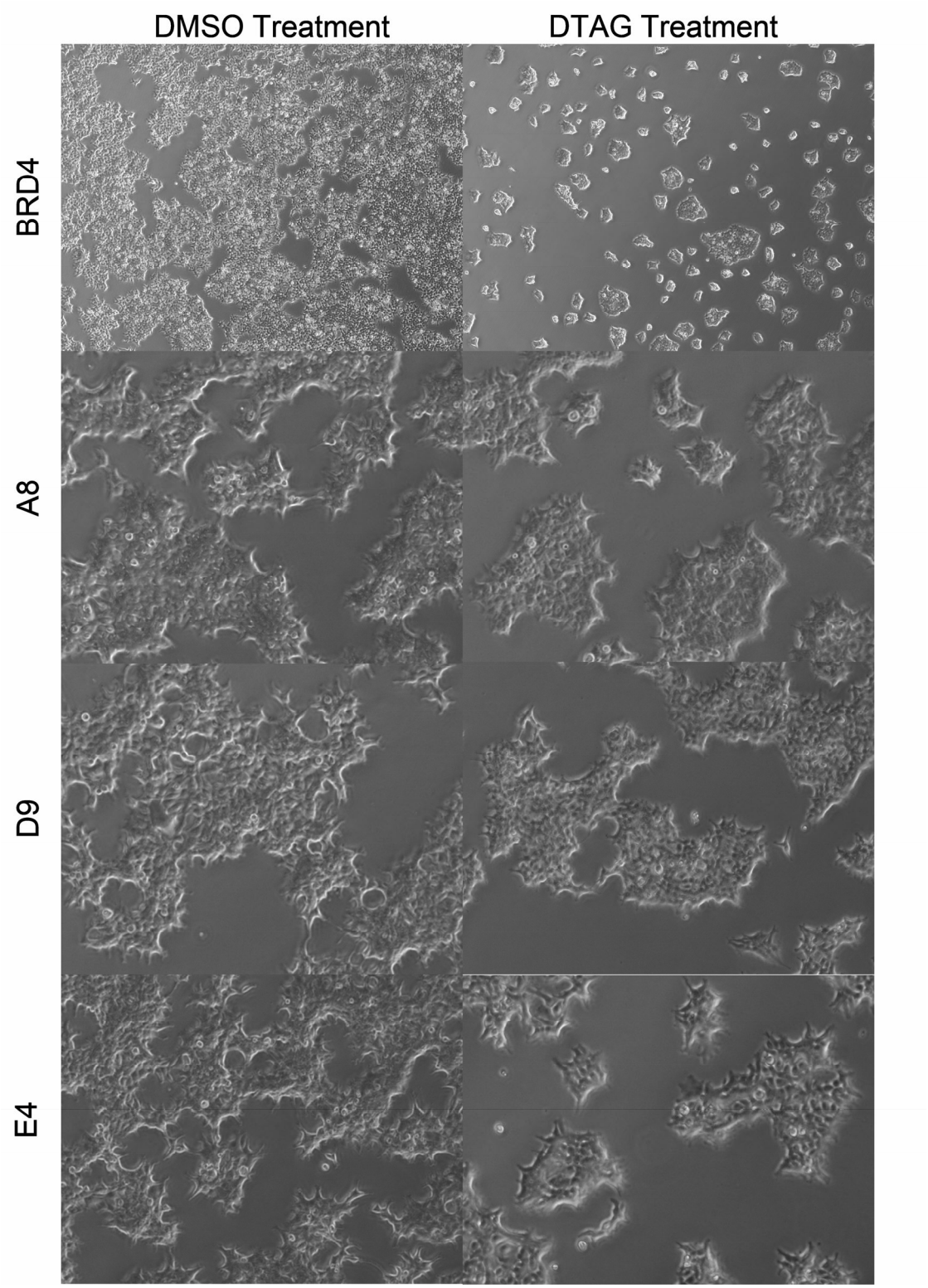


Figure 12: Confluency and morphological effects of dTAG-13. Homozygous BRD4-tagged cells treated with dTAG-13 display a significant decrease in confluency and a far smooth, self-contained colony morphology. Heterozygous SETD5-tagged colonies show a more subtle morphological change whereby colonies have smoother edges.

Validation with HA-tag Affirms Partial SETD5 Degradation

Afterwards, we sought to validate the degradation of SETD5 by performing a dTAG experiment and blotting for the HA-tag. The results showed complete degradation of the HA-tag for the *BRD4*-tagged cells treated with dTAG-13, and partial degradation of the HA-tag for the *SETD5*-tagged A8 line treated with dTAG-13. Stripping and re-blotting of the membrane for SETD5 demonstrated that the HA-tag on the *SETD5*-tagged cells overlaps with the lower SETD5 band. This reaffirms the findings in the prior blot whereby the lower SETD5 band displayed greater depletion. The SETD5 re-blot demonstrated a comparable amount of degradation of the dTAG-13 treated *SETD5*-tagged cells. The HEK293T Wild type negative control displayed no result for the HA-tag blot as expected. This suggests that the HA-tag and SETD5 probing results are consistent with one another in demonstrating SETD5 degradation.

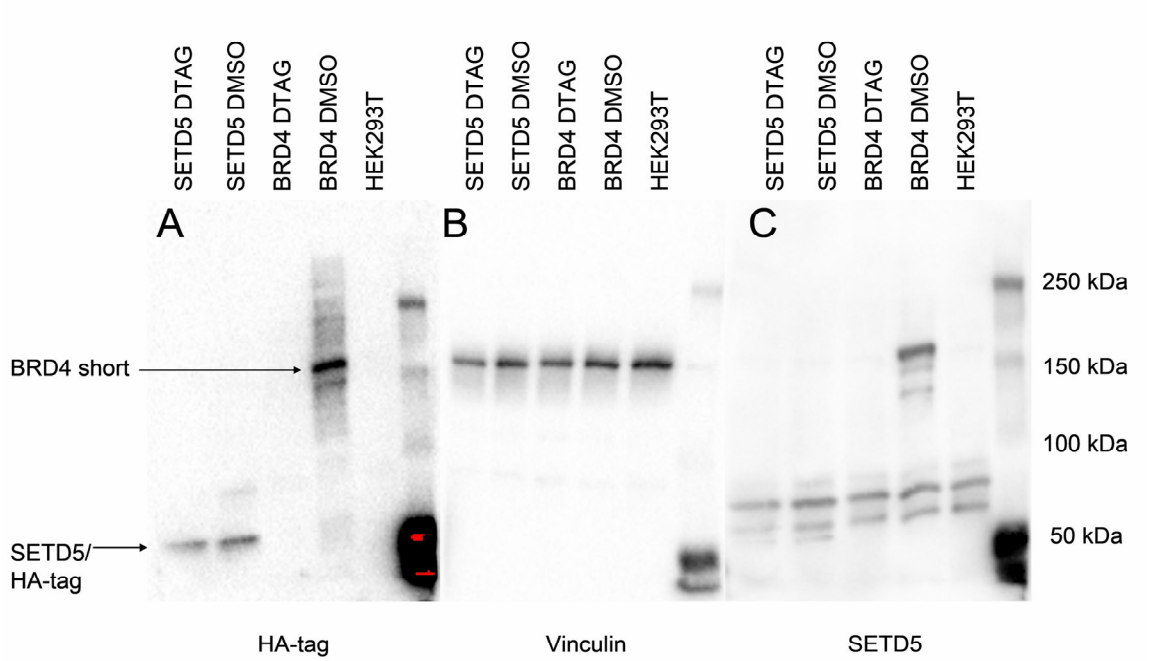


Figure 13: HA-tag Validation of SETD5 Degradation. (A) Complete degradation of HA-tag in homozygous BRD4-tagged cells compared to partial degradation of HA-tag in heterozygous SETD5-tagged cell line (A8 clone). HEK293 WT negative control shows no expression of HA-tag. (B) Vinculin loading control demonstrates even loading of proteins. (C). SETD5 re-blot of stripped HA-tag membrane confirms the results of the HA-tag membrane and also confirms suspicion that the lower band of SETD5 band pattern is the isoform demonstrating degradation.

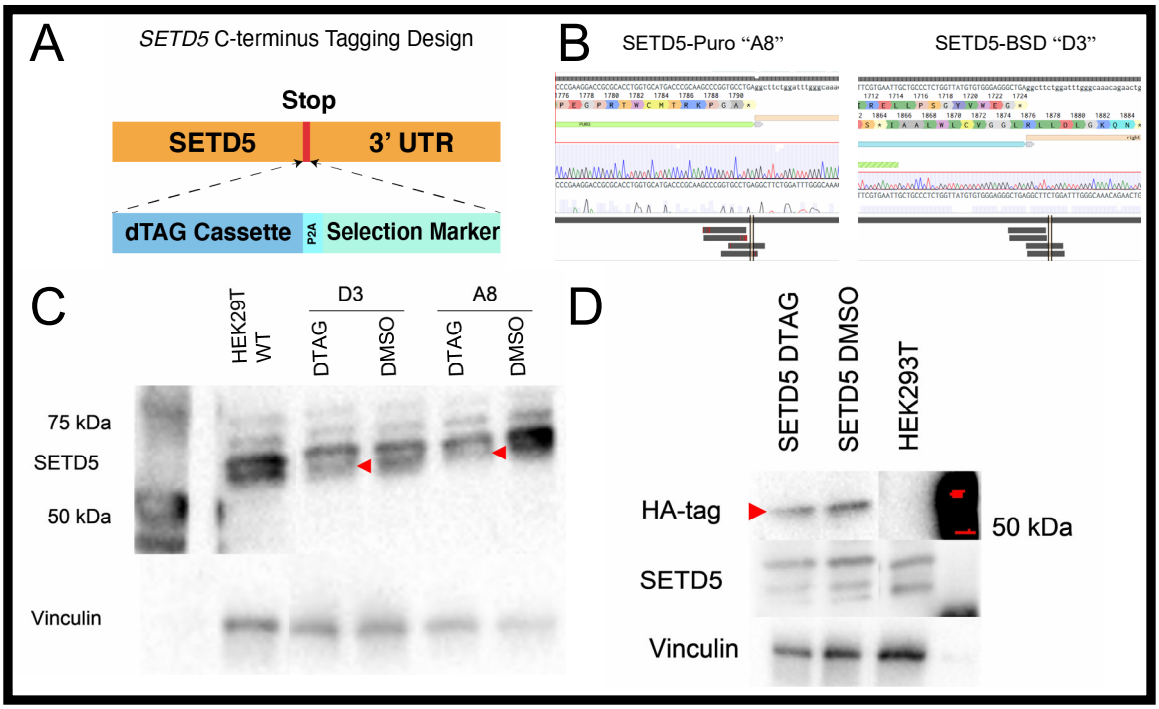


Figure 14: Graphical summary of tagging and depletion of SETD5. (A) Knock-in design of dTAG cassette at the *SETD5* Stop codon. The dTAG cassette polycistronically encodes the tag and a selection marker. (B) Sequencing alignment results of two confirmed HEK293T SETD5-tagged clones, A8 and D3. (C) Partial degradation of heterozygous SETD5 protein in clones A8 and D3. (D) Validation of SETD5 tagging and partial degradation by blotting for HA-tag.

DISCUSSION

In this paper we described current studies in the field of Autism research and the significance that Chromatin Regulators play within the etiology of ASD. We then constructed CRISPR and HDR plasmid vectors for the tagging of the Chromatin Regulator *SETD5* at the Stop codon site for inducible protein depletion. We confirmed the sequence of two *SETD5*-tagged heterozygotes and demonstrated partial depletion of the SETD5 protein in a dTAG experiment followed by western blot. Next we confirmed this degradation through blotting of the HA-tag on the SETD5 protein, which demonstrated partial degradation relative to BRD4-tagged homozygous cells demonstrating complete degradation.

Continuing this work, the next steps would be to test the reversibility of the dTAG system in the SETD5-tagged clones. In our preliminary tests for reversibility of the dTAG system in BRD4-tagged cells, we observed reversibility but the results were partially inconclusive due antibody recognition when the protocols of the western blot were still being established. Yet the original source of the dTAG system indicated that reversal of the protein depletion is possible with transferal of cells to a different plate. Washouts are not sufficient in 2D cultures, as the dTAG-13 ligand is dissolved in DMSO which tends to remain on the surface of the plate. Therefore, optimizations must be made to achieve reversibility in the organoid model. For 3D culture these optimizations can include experimenting with washouts, passaging, or manual picking of organoids to transfer to a new plate with new media.

In addition, ATAC-seq and ChIP-seq should be performed on untreated, DMSO-treated, and dTAG-13 treated SETD5-tagged cells alongside HEK293T WT cells in order

to assess the effects that tagging has on the SETD5 protein product, and furthermore how depletion affects expression and chromatin architecture.

These plasmid vectors for the tagging of the SETD5 gene provide a novel means of studying genes involved in neurodevelopmental disorders, or more specifically chromatin regulators involved in ASD. The advantage to this system is the flexibility with which one can induce rapid degradation at any time point, which can help with better recapitulation of post-zygotic mutations that occur along neurodevelopment. In addition, the ability to reverse this degradation will allow for higher contrast between the wild type and haploinsufficient states. Overall, our framework for the introduction of an inducible protein depletion tag to the *SETD5* gene provides us with an advantageous means of studying *SETD5* along neurodevelopment, and provides a model for the study of other CRs and genes as well.

REFERENCES

1. The National Institute of Mental Health. Autism Spectrum Disorder. https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml. Updated March 2018. Accessed April 2020.

2. Ha S, Sohn IJ, Kim N, Sim HJ, Cheon KA. Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan. *Exp Neurobiol*. 2015;24(4):273-284. doi:10.5607/en.2015.24.4.273

3. Autsim Statistics and Facts. Autism Speaks. https://www.autismspeaks.org/autism-statistics. Accessed May 17, 2020.

4. Pan YH, Wu N, Yuan XB. Toward a Better Understanding of Neuronal Migration Deficits in Autism Spectrum Disorders. *Front Cell Dev Biol*. 2019;7:205. Published 2019 Sep 20. doi:10.3389/fcell.2019.00205

5. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Mahajan M, Manaa D, Pawitan Y, Reichert J, Ripke S, Sandin S, Sklar P, Svantesson O, Reichenberg A, Hultman CM, Devlin B, Roeder K, Buxbaum JD. Most genetic risk for autism resides with common variation. *Nat Genet*. 2014;46(8):881-885. doi:10.1038/ng.3039

6. SFARI Gene. SFARI. https://www.sfari.org/resource/sfari-gene. Published February 27, 2020. Accessed February, 2019.

7. Tau GZ, Peterson BS. Normal development of brain circuits. *Neuropsychopharmacology*. 2010;35(1):147-168. doi:10.1038/npp.2009.115

8. Gabriele, M., Lopez Tobon, A., D'Agostino, G. and Testa, G., 2018. The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 84, pp.306-327.

9. Pescosolido MF, Yang U, Sabbagh M, Morrow EM. Lighting a path: genetic studies pinpoint neurodevelopmental mechanisms in autism and related disorders. *Dialogues Clin Neurosci*. 2012;14(3):239-252.

10. Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*. 2014;15(2):93-106. doi:10.1038/nrg3607

11. Lasalle JM. Autism genes keep turning up chromatin. *OA Autism*. 2013;1(2):14. doi:10.13172/2052-7810-1-2-610

12. Suliman R, Ben-David E, Shifman S. Chromatin regulators, phenotypic robustness, and autism risk. *Front Genet*. 2014;5:81. Published 2014 Apr 10. doi:10.3389/fgene.2014.00081

13. Lino CA, Harper JC, Carney JP, Timlin JA. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv*. 2018;25(1):1234-1257. doi:10.1080/10717544.2018.1474964

14. Zimmer AM, Pan YK, Chandrapalan T, Kwong RWM, Perry SF. Loss-of-function approaches in comparative physiology: is there a future for knockdown experiments in the era of genome editing? *The Journal of Experimental Biology*. 2019;222(7). doi:10.1242/jeb.175737

15. Larson M, Gilbert L, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8,** 2180–2196 (2013). https://doi.org/10.1038/nprot.2013.132

16. Sathyan, K., McKenna, B., Anderson, W., Duarte, F., Core, L., and Geurtin, M., 2019. An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion. *Genes & Development*, 33(19-20), pp.1441-1455.

17. Nabet B, Roberts JM, Buckley DL, Paulk J, Dastjerdi S, Yang A, Leggett A, Erb MA, Lawlor MA, Souza A, Scott TG, Vittori S, Perry JA, Jun Q, Winter GE, Wong KK, Gray NS, Bradner JE. The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol*. 2018;14(5):431-441. doi:10.1038/s41589-018-0021-8

18. Behnam Nabet. dTAG – You're it!. https://blog.addgene.org/dtag-youre-it. Published June 21, 2018. Accessed May 17, 2020.

19. Van Dijck A, Vulto-van Silfhout AT, Cappuyns E, van der Werf IM, Mancini GM, Tzschach A, Bernier R, Gozes I, Eichler EE, Romano C, Lindstrand A, Nordgren A, Kvarnung M, Kleefstra T, de Vries BBA, Kury S, Rosenfeld JA, Meuwissen ME, Vandeweyer G, Kooy RF. Clinical Presentation of a Complex Neurodevelopmental Disorder Caused by Mutations in ADNP. *Biol Psychiatry*. 2019;85(4):287-297. doi:10.1016/j.biopsych.2018.02.1173

20. van der Sluijs PJ, Jansen S, Vergano SA, Adachi-Fukuda M, Alanay Y, AlKindy A, Baban A, Bayat A, Beck-Wödl S, Berry K, Bijlsma EK, Bok LA, Brouwer AFJ, van der Burgt I, Campeau PM, Canham N, Chrzanowska K, Chu YWY, Chung BHY, Dahan K, De Rademaeker M, Destree A, Dudding-Byth T, Earl R, Elcioglu N, Elias ER, Fagerberg C, Gardham A, Gener B, Gerkes EH, Grasshoff U, van Haeringen A, Heitink KR, Herkert JC, den Hollander NS, Horn D, Hunt D, Kant SG, Kato M, Kayserili H, Kersseboom R, Kilic E, Krajewska-Walasek M, Lammers K, Laulund LW, Lederer D, Lees M, López-González V, Maas S, Mancini GMS, Marcelis C, Martinez F, Maystadt I, McGuire M, McKee S, Mehta S, Metcalfe K, Milunsky J, Mizuno S, Moeschler JB, Netzer C, Ockeloen CW, Oehl-Jaschkowitz B, Okamoto N, Olminkhof SNM, Orellana C, Pasquier L, Pottinger C, Riehmer V, Robertson SP, Roifman M, Rooryck C, Ropers FG, Rosello M, Ruivenkamp CAL, Sagiroglu MS, Sallevelt SCEH, Sanchis Calvo A, Simsek-

Kiper PO, Soares G, Solaeche L, Sonmez FM, Splitt M, Steenbeek D, Stegmann APA, Stumpel CTRM, Tanabe S, Uctepe E, Utine GE, Veenstra-Knol HE, Venkateswaran S, Vilain C, Vincent-Delorme C, Vulto-van Silfhout AT, Wheeler P, Wilson GN, Wilson LC, Wollnik B, Kosho T, Wieczorek D, Eichler E, Pfundt R, de Vries BBA, Clayton-Smith J, Santen GWE. The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to Coffin-Siris syndrome [published correction appears in Genet Med. 2019 Jan 29;:]. *Genet Med*. 2019;21(6):1295-1307. doi:10.1038/s41436-018-0330-z

21. Balasubramanian M, Willoughby J, Fry AE, Weber A, Firth HV, Deshpande C, Berg JN, Chandler K, Metcalfe KA, Lam W, Pilz DT, Tomkins S. Delineating the phenotypic spectrum of Bainbridge-Ropers syndrome: 12 new patients withde novo, heterozygous, loss-of-function mutations inASXL3and review of published literature. *Journal of Medical Genetics*. 2017;54(8):537-543. doi:10.1136/jmedgenet-2016-104360

22. Carvill G, Helbig I, Mefford H. CHD2-Related Neurodevelopmental Disorders. 2015 Dec 10. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2020. Available from: https://www.ncbi.nlm.nih.gov/books/NBK333201/

23. Poisson A, Chatron N, Labalme A, Fourneret P, Ville D, Mathieu ML, Sanlaville D, Demily C, Lesca G. Chromatin remodeling dysfunction extends the etiological spectrum of schizophrenia: a case report. *BMC Med Genet*. 2020;21(1):10. Published 2020 Jan 8. doi:10.1186/s12881-019-0946-0

24. Barnard RA, Pomaville MB, O'Roak BJ. Mutations and Modeling of the Chromatin Remodeler CHD8 Define an Emerging Autism Etiology. *Front Neurosci*. 2015;9:477. Published 2015 Dec 17. doi:10.3389/fnins.2015.00477

25. An Y., Zhang L., Liu W, Jiang Y, Chen X, Lan X, Li G, Hang Q, Wang J, Gusella JF, Du Y, Shen Y. De novo variants in the Helicase-C domain of CHD8 are associated with severe phenotypes including autism, language disability and overgrowth. *Hum Genet* **139,** 499–512 (2020). https://doi.org/10.1007/s00439-020-02115-9

26. Feldman, HR, Dlouhy, SR, Lah, MD, Payne, KK, Weaver, DD. The progression of Wiedemann–Steiner syndrome in adulthood and two novel variants in the *KMT2A* gene. *Am J Med Genet Part A*. 2019; 179A: 300– 305. https://doi.org/10.1002/ajmg.a.60698

27. White J, Beck CR, Harel T, Posey JE, Jhangiani SN, Tang S, Farwell KD, Powis Z, Mendelsohn NJ, Baker JA, Pollack L, Mason KJ, Wierenga KJ, Arrington DK, Hall M, Psychgios A, Fairbrother L, Walkiewicz M, Person RE, Niu Z, Zhang J, Rosenfeld JA, Muzny DM, Eng C, Beaudet AL, Lupski JR, Boerwinkle E, Gibbs RA, Yang Y, Xia F, Sutton VR . POGZ truncating alleles cause syndromic intellectual disability. *Genome Med*. 2016;8(1):3. Published 2016 Jan 6. doi:10.1186/s13073-015-0253-0

28. Fernandes, I.R., Cruz, A.C., Ferrasa, A., Phan, D., Herai, R.H. and Muotri, A.R. (2018), Genetic variations on *SETD5* underlying autistic conditions. Devel Neurobio, 78: 500-518. doi:10.1002/dneu.22584

29. Z Powis, K D Farwell Hagman, C Mroske, K McWalter, J S Cohen, R Colombo, A Serretti, A Fatemi, K L David, J Reynolds, L Immken, H Nagakura, C M Cunniff, K Payne, T Barbaro-Dieber, K W Gripp, L Baker, T Stamper, K A Aleck, E S Jordan, J H Hersh, J Burton, I M Wentzensen, M J Guillen Sacoto, R Willaert, M T Cho, I Petrik, R Huether, S Tang . Expansion and further delineation of the *SETD5* phenotype leading to global developmental delay, variable dysmorphic features, and reduced penetrance. *Clin Genet*. 2018;93:752–761. https://doi.org/10.1111/cge.13132

30. Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M., Son, J., Reinberg, D., Lachner, M. and Jenuwein, T., 2012. Prdm3 and Prdm16 are H3K9me1 Methyltransferases Required for Mammalian Heterochromatin Integrity. *Cell*, 150(5), pp.948-960.

31. Deliu E, Arecco N, Morandell J, Dotter CP, Contreras X, Girardot C, Käsper EL, Kozlova A, Kishi K, Chiaradia I, Noh KM, Novarino G.. Haploinsufficiency of the intellectual disability gene *SETD5* disturbs developmental gene expression and cognition. *Nat Neurosci* **21,** 1717–1727 (2018). https://doi.org/10.1038/s41593-018-0266-2

32. Kuechler A, Zink AM, Wieland T, Lüdecke HJ, Cremer K, Salviati L, Magini P, Najafi K, Zweier C, Czeschik JC, Aretz S, Endele S, Tamburrino F, Pinato C, Clementi M, Gundlach J, Maylahn C, Mazzanti L, Wohlleber E, Schwarzmayr T, Kariminejad R, Schlessinger A, Wieczorek D, Strom TM, Novarino G, Engels H. Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur J Hum Genet*. 2015;23(6):753-760. doi:10.1038/ejhg.2014.165

33. Grozeva D, Carss K, Spasic-Boskovic O, Parker MJ, Archer H, Firth HV, Park SM, Canham N, Holder SE, Wilson M, Hackett A, Field M, Floyd JA; UK10K Consortium, Hurles M, Raymond FL. De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am J Hum Genet*. 2014;94(4):618-624. doi:10.1016/j.ajhg.2014.03.006

34. Fang YL, Zhang RP, Wang YZ, Cao LR, Zhang YQ, Cai CQ. A novel mutation in a common pathogenic gene (*SETD5*) associated with intellectual disability: A case report. *Exp Ther Med*. 2019;18(5):3737-3740. doi:10.3892/etm.2019.8059

35. Moore SM, Seidman JS, Ellegood J, Gao R, Savchenko A, Troutman TD, Abe Y, Stender J, Lee D, Wang S, Voytek B, Lerch JP, Suh H, Glass CK, Muotri AR. *Setd5* haploinsufficiency alters neuronal network connectivity and leads to autistic-like behaviors in mice. *Transl Psychiatry* **9,** 24 (2019). https://doi.org/10.1038/s41398-018-0344-y

36. Osipovich AB, Gangula R, Vianna PG, Magnuson MA. Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation. *Development*. 2016;143(24):4595-4607. doi:10.1242/dev.141465

37. HUGO Gene Nomenclature Committee. HGNC. https://www.genenames.org/download/statistics-and-files/. Accessed May 17, 2020.

38. Epifactors Database. Epifactors Database. https://epifactors.autosome.ru/. Accessed May 17, 2020.

39. Sakuma T, Nakade S, Sakane Y, Suzuki KT, Yamamoto T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc*. 2016;11(1):118-133. doi:10.1038/nprot.2015.140